Fine-grained Sleep Monitoring: Hearing Your Breathing with Smartphones

Yanzhi Ren[1], Chen Wang[1], Jie Yang[2], Yingying Chen[1]
[1]Department of ECE, Stevens Institute of Technology, Hoboken, NJ 07030
{yren2, cwang42, yingying.chen}@stevens.edu
[2]Department of CS, Florida State University, Tallahassee, FL 32306
jyang5@fsu.edu

***Abstract*—Sleep monitoring has drawn increasingly attention as the quality and quantity of the sleep are important to maintain a person's health and well-being. For example, inadequate and irregular sleep are usually associated with serious health problems such as fatigue, depression and cardiovascular disease. Traditional sleep monitoring systems, such as PSG, involve wearable sensors with professional installations, and thus are limited to clinical usage. Recent work in using smartphone sensors for sleep monitoring can detect several events related to sleep, such as body movement, cough and snore. Such coarse-grained sleep monitoring however is unable to detect the breathing rate which is an important vital sign and health indicator. This work presents a fine-grained sleep monitoring system which is capable of detecting the breathing rate by leveraging smartphones. Our system exploits the readily available smartphone earphone placed close to the user to reliably capture the human breathing sound. Given the captured acoustic sound, our system performs noise reduction to remove environmental noise and then identifies the breathing rate based on the signal envelope detection. Our system can further detect detailed sleep events including snore, cough, turn over and get up based on the acoustic features extracted from the acoustic sound. Our experimental evaluation of six subjects over six months time period demonstrates that the breathing rate monitoring and sleep events detection are highly accurate and robust under various environments. By combining breathing rate and sleep events, our system can provide continuous and noninvasive fine-grained sleep monitoring for healthcare related applications, such as sleep apnea monitoring as evidenced by our experimental study.**

I. Introduction

There exists a broad array of healthcare related applications that benefit from fine-grained sleep monitoring – non-obtrusive breathing rate monitoring for the understanding of the sleep quality. For example, inadequate and irregular sleep can lead to serious health problems such as fatigue, depression, cardiovascular disease and anxiety [1]. And breathing rate monitoring is critical to detect early signs of several diseases such as diabetes and heart disease [2], [3]. The breathing rate monitoring can also be applied in the sleep apnea diagnosis and treatment [4], treatment for asthma [5] and sleep stage detection [6]. It is thus important to enable the fine-grained sleep monitoring to facilitate these healthcare related applications.

The challenge in fine-grained sleep monitoring lies in solutions providing breathing rate detection without requiring invasive diagnostic devices and at a minimal cost. Traditional sleep monitoring systems, such as Poly-somnography (PSG) [7], are able to provide fine-grained sleep monitoring including breathing rate detection. These systems however involve multiple wearable sensors and professional installations, and thus are limited to clinical usage. The high complexity and cost prevent these systems from large scale and long term deployment. Some commercial monitoring products including *ZEO*, *Fitbit* and *Sleep Tracker* [8] have been designed based on PSG. They require the user's involvement and are hard to use: the user is required to wear a device during sleep, which may affect or change his/her daily sleep habits. Several smartphone Apps, such as *Sleep as Android*, *Sleep cycle alarm clock* and *iSleep* [9], use the smartphone's built-in microphone and motion sensors to perform noninvasive sleep monitoring. These apps can only support coarse-grained monitoring such as the detection of body movements, coughing and snoring.

This work demonstrates that it is possible to achieve fine-grained non-invasive sleep monitoring at a minimal cost without the involvement of diagnostic devices by exploiting the off-the-shelf smartphone and its earphone. Our system captures the breathing sound generated by the air flow for breathing rate detection. Studies [10] show that the airway flow is correlated with the amplitude of the respiratory sound during normal breathing. It is thus possible to monitor the breathing rate based on the breathing sound. Several unique challenges present when utilizing the breathing sound for supporting the fine-grained sleep monitoring. First, the sound of the user's breathing is usually weak, posing difficulty in capturing the breathing sound by using the smartphone's built-in microphone. Second, the background noise generated by various sources including the air conditioning, heater and traffic outside the house significantly affects the recording of the breathing sound. Third, the acoustic characteristics of the breathing sound vary dramatically among users, requiring the breathing rate detection method to be adaptive to different users.

Specifically, we take the viewpoint by using the low-cost smartphone's earphone to capture human breathing sound. The earphone, a standard piece in phone's sales package, has been widely used when listening to music, watching videos and making calls on smartphones. A recent study [11] reports that over 70 percent of smartphone users listen to music or make calls daily using earphones. Furthermore, before going to sleep at night, people tend to use earphones when listening to music, news, stories, etc. from smart devices and place the earphone aside on the pillow when sleeping. They can also intentionally place earphones in such positions for monitoring purposes if needed. This is different from asking users to place the smartphone close to them during sleep. It is thus natural to exploit the earphone to capture the breathing sound. We find that the sound monitoring quality could be largely improved

978-1-4799-8381-0/15/$31.00 ©2015 IEEE

when using the earphone (although it is a small add-on piece to the smartphone) comparing using the standard built-in sensors on smartphones.

The benefit of using earphone is four-fold. First, the microphone on earphone has a higher recording quality than that of the smartphone built-in microphone, resulting in more reliable recorded breathing sound. Second, many users are resistant to place the smartphone close to them during sleep due to the concern of electromagnetic radiation. However, users tend to leave the earphone plugged into their ears or put it aside on their pillows during sleep. Third, we find that the earbuds on earphone could be used as microphones, which helps to enhance the recording ability of the breathing sound. Fourth, using earphone can also capture other sleep related events easily such as snoring, coughing, turn-over and get up.

Given the recorded acoustic sound from the earphone, we perform noise estimation and subtraction to reduce the impact of background noise. We then exploit the high correlation between a user's breathing cycles to make our breathing rate detection method adaptive to different users. Finally, we use acoustic features extracted from acoustic sound for sleep event (e.g., snoring, coughing, turn-over and get up) detection. By combining breathing rate and sleep events, our system provides continuous and noninvasive fine-grained sleep monitoring for healthcare related applications (such as sleep apnea monitoring) as evidenced by our experimental study. We summarize our main contributions as follows:

- We find that using the smartphone earphone is a promising approach to capture the breathing rate for fine-grained sleep monitoring.
- Our system built on the smartphone earphone can achieve continuous and noninvasive breathing rate monitoring without involving additional diagnostic devices.
- By exploiting the correlation relationship inherent in a user's breathing cycles, our breathing rate detection method is adaptive to different users.
- Our approach has the capability to detect various sleep events easily including snoring, coughing, turn over and get up. By combining the detected breathing rate and sleep events, our system facilitates healthcare related applications, such as sleep apnea monitoring.
- We evaluate our system with six subjects over six months time period. The results show that our system is highly accurate and robust in breathing rate monitoring and sleep event detection under various scenarios.

The rest of the paper is organized as follows. We first present the related studies in Section II. We then describe the design of our fine-grained sleep monitoring system in Section III. Next, we present the system implementation and case study in Section IV. In Section V, we validate the feasibility of our proposed system through real experiments. Finally, we conclude our work in Section VI.

II. Related Work

Medical-based sleep monitoring systems are developed for clinical usage. In particular, Polysomnography (PSG) [7] is used in medical facilities to perform fine-grained sleep monitoring by attaching multiple sensors on patients, which require professional installations. It can measure many body functions during sleep including breathing functions, eye movements (EOG), heart rhythm (ECG) and muscle activity. Such systems incur high cost and are usually limited to clinical usage. Actigraphy [12] has been developed as an affordable alternative to monitor human sleep and wakefulness based on the body movement detection. The Actigraphy however cannot monitor breathing rate.

Furthermore, light-weight coarse-grained sleep monitoring products are developed based on PSG or Actigraphy. A popular monitoring product called ZEO measures the electrical signals produced by the brain using sensors in the headband worn by users to infer sleep patterns. Fitbit and Sleep Tracker [8] detect body movements by using wearable accelerometer sensor to infer how long and how well the person sleep. These products mainly infer the people's sleep length and do not have the capability to perform fine-grained sleep monitoring. They also require user's involvement by wearing a device during sleep, which may affect the user's sleep habits, and many are even resistant to follow.

There are also recent work dedicated for breathing rate monitoring. The capnometer system [13] uses the gas analyzer to measure the carbon dioxide concentration in exhaled air. However, the price is relatively high and may not be affordable for some users. The recent work of using wireless network [14] for detecting breathing rate has the limitation of high cost as well: dedicated wireless sensors are used to monitor the changes of received signal caused by breathing.

Several smartphone Apps, such as *Sleep as Android*, *Sleep cycle alarm clock* and *iSleep* [9], can perform low cost sleep monitoring by using the smartphone built-in microphone and motion sensors. These apps however only provide coarse-grained monitoring such as the detection of body movements, coughing and snoring. [15] utilizes the phone usage features such as the duration of phone lock to measure sleep duration. The *Respiratory* app [16] derives a person's respiratory rate by analyzing the movements of the user's abdomen when placing the phone between the user's ribcage and stomach. This method requires user's involvement and attention, and is thus hard for continuous fine-grained sleep monitoring.

Our work is different in that we provide fine-grained sleep monitoring by leveraging the smartphone earphone, which is readily available sold together with almost all the phones. Our system is low-cost, nonobtrusive and easy-to-use without requiring dedicated sensors or professional installation.

III. System Design

In this section, we discuss the system requirements and provide an overview of our system design.

A. System Requirements

Our system aims to provide continuous and noninvasive fine-grained sleep monitoring using smartphones. Specifically, our system is designed to meet the following requirements:

Supporting Fine-grained Sleep Monitoring. Our system should be able to detect both breathing rate and detailed sleep events. Such fine-grained monitoring is essential for many healthcare related applications, such as the diagnosis of sleep apnea, asthma, diabetes and heart disease.

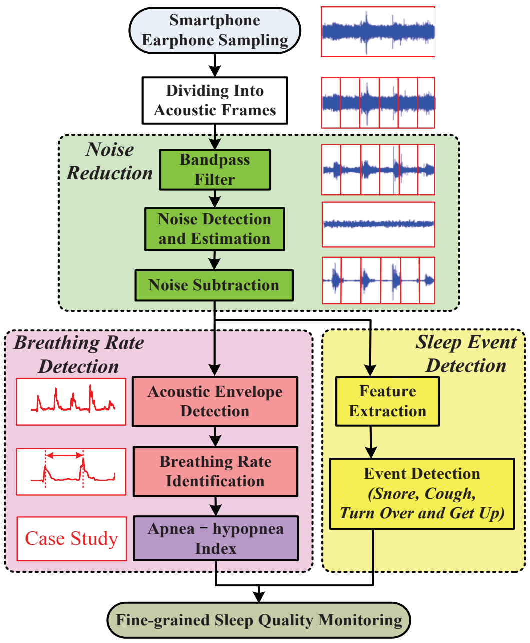


Fig. 1. System flow of fine-graind sleep monitoring.

Easy to Use. Since our system is designed to perform continuous sleep monitoring, it should be noninvasive: no sensors are needed to be attached to the user as the wearable sensors may affect the user's sleep habits. Moreover, our system should be low-cost: re-use existing devices in our daily life without professional installation.

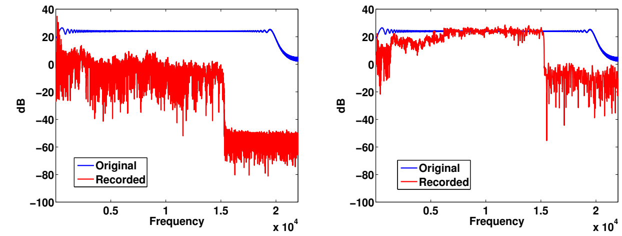
Robust Across Different Environments. The background noise generated by various sources including the air conditioning, heater and traffic outside the house is unavoidable during sleep. Our system should be able to provide effective fine-grained sleep monitoring under such noisy environments.

Light Weight. Due to the limited computational resources of the smartphones, the designed algorithms should be lightweight in order to process the collected data in real time.

B. System Overview

The basic idea of our system is to use smartphone's earphone to capture the breathing sound for fine-grained sleep monitoring. The earphone is included in most off-the-shelf smartphones' sales package as a standard accessory and it is used with smartphones extensively: over 70 percent of smartphone users like to carry earphones with their smartphones while watching videos, enjoying music and making calls [11]. Furthermore, many people tend to use earphones to listen to music, news, or other programs from smart devices before they fall asleep. It is thus possible for us to explore utilizing the smartphone's earphone to capture breathing sound for sleep monitoring.

As illustrated in Figure 1, the system takes as input the time-series acoustic signal captured by the earphone placed close to the user. The sound can be further enhanced by using the input from the earbuds after connecting the earbuds with the output of the microphone by using off-the-shelf connecters. The acoustic signal is then preprocessed to remove environmental noise via noise reduction. The next two components of our system are the breathing rate detection and sleep event detection. Given the input sound, we first perform signal envelop detection, based on which the breathing rate can be identified adaptively for different users. In the meanwhile, our system also has the ability to detect the sound of sleep related events, such as snoring, coughing, turn over and get up. And the sleep event detection can be done based on the extracted acoustic features from the recorded sound. Such capabilities enable our fine-grained sleep monitoring system to support many healthcare related applications in the way the coarse-grained monitoring approaches not possible, such as the diagnosis of sleep apnea, asthma, diabetes and heart disease.



(a) (b)

Fig. 2. Frequency response of (a) Smartphone built-in microphone and (b) Microphone on earphone

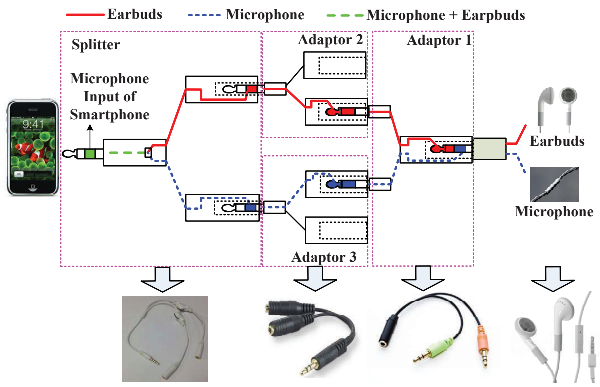
To demonstrate the usability of our system, we perform a case study of sleep apnea monitoring. In particular, we calculate the Apnea-hypopnea Index based on the detected breathing rate for sleep apnea monitoring in Section IV-D.

C. The Use of Earphone

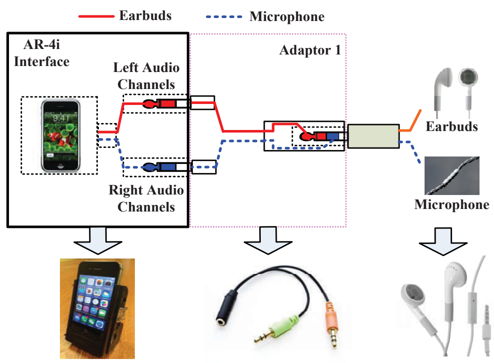
Although earphone is a small add on piece to smartphones, our system benefits from the earphone in several aspects: high recording quality when comparing to smartphone built-in microphone; small size and extension wires make the earphone reach out to the user; the earbuds could be utilized for enhancing the sound quality; and the capability of stereo recording. We next introduce several ways the earphone can be used to record breathing sound. Specifically, we can fully utilize the *microphone* on earphone, the *earbuds* and the *stereo recording* capability.

1) Microphone on earphone: We find that the microphone on the earphone has a better recording quality than the built-in microphone on the smartphone. It is also relatively easier to keep the earphone close to user because of its small size and extension wires. These advantages make the earphone's microphone a more reliable source to record the weak breathing sound during sleep.

Figure 2 depicts the recorded sound quality comparison between the earphone's microphone and the smartphone built-in microphone. In particular, Figure 2 (a) and (b) show the spectrum of sounds recorded by the built-in microphone on Iphone 4 and the earphone's microphone when we play a chirp signal with 50 milliseconds from 0 Hz to 20000 Hz, respectively. We place the recording devices at the same distance from the speaker. We find that the spectrum of sounds recorded by the earphone's microphone is very similar to the spectrum of the original chirp signal, while the spectrum of sounds recorded by the built-in microphone degrades significantly from that of the original chirp signal. This indicates that the earphone's microphone has better frequency response than the built-in microphone, suggesting the earphone's microphone



(a) Earbuds for mono recording



(b) Earbuds for stereo recroding

Fig. 3. Illustration of using earbuds for recording.

is a better choice for sound recording in addition to its advantages of small size and extension wires.

2) Earbuds: The earbuds on earphone are essentially microphones. Earbuds produce acoustic waves, and they can also pick up sound in the same way as microphone does. Earbuds and microphones are both transducers and they can perform the same functionality. We thus can utilize earbuds as additional microphones to enhance the sound quality.

Figure 3 (a) shows the detailed connections when we use the earbuds to capture the breathing sound and connect the output of earbuds with the phone's microphone. Specifically, the connections consist of one earphone splitter and three earphone adapters: the splitter can split one set of audio output/input to two separate sets of output/input. Thus, it allows us to connect two separate microphone connections together for use with the smartphone. Whereas the Adaptor 1 converts the earphone plug into two plugs: one for earbuds and the other for the earphone's microphone. Adaptor 2 and 3 then enable these two plugs to be connected with the two microphone outputs of the splitter. Finally, the splitter combines the plug for earbuds and plug for the earphone's microphone together and connect it with the microphone output on the smartphone. The earbuds then work as additional microphones during the recording to improve the sound quality.

3) Stereo recording: Since the earbuds can be used for recording, we can further perform stereo recording (i.e., two channels) with one single earphone: the earbuds as the left stereo input channel and the microphone as the right stereo input channel. To enhance the recording quality, we then combine two audio channels by deriving the maximum samples from these two independent inputs, rather than processing samples from each channel separately.

Figure 3 (b) illustrates the stereo audio recording with one single earphone by using an audio interface (Fostex AR-4i [17]) and an adapter. The audio interface offers a stereo audio recording and is increasingly popular in recent years as more and more people use it together with their smartphones. Similarly, the Adaptor 1 also converts the earphone plug into two plugs: one for earbuds and the other for the earphone's microphone. The plug for earbuds is then connected to the left stereo input channel, whereas the plug for earphone's microphone is connected to the right stereo input channel of AR-4i. We can therefore obtain more reliable breathing sound under stereo recording, especially under the noisy environments.

In our system implementation, we evaluate our system with different usage of the earphone. The following three usages have been implemented based on the availability of connector and audio interface. 1) *Microphone Only*: only the microphone on earphone is used for recording; 2) *Earbuds + Microphone*: both the microphone and earbuds are used for mono recording; 3) *Stereo Recording*: two earbuds are used to record as one independent channel and the microphone is used to record as another independent channel. The detailed performance evaluation of these three recording methods is presented in Section V.

IV. System Implementation and Case Study

In this section, we present the detailed system implementation of our fine-grained sleep monitoring system.

A. Noise Reduction

Most people usually sleep in a relatively quiet environment. However, the background noise generated by various sources including the air conditioning, heater and traffic outside the house is unavoidable during sleep. The thermal noise from recording equipments while recording also affects the breathing sound recording. Both the background and thermal noise acoustically added to the breathing sound may significantly degrade the performance of sleep monitoring. To build a robust system, we first perform noise reduction to reduce the effects of the the background and thermal noise.

1) Noise Detection: Noise reduction aims to clean the recorded acoustic sounds by subtracting the estimated noise components. In our work, we propose to estimate the noise components from time frames that only contain ambient noise (i.e., noise frames). We thus first distinguish the noise frames from the non-noise frames in the recorded acoustic sounds.

Specifically, the collected acoustic signal is first sampled with the frequency of 8 kHz and segmented into frames with $K = 800$ samples each. A bandpass filter is then applied to the raw audio data to remove both high and low frequency sounds that are not related to the breathing events. The lower and upper cutoff frequencies are set as 100 Hz to 3400 Hz. We choose the lower cutoff frequency as 100 Hz because it can remove most of the common electronic noise at lower frequency band. The upper cutoff frequency at 3400 Hz is selected because the most sound generated by breathing or sleep related events is below 3400 Hz [18].

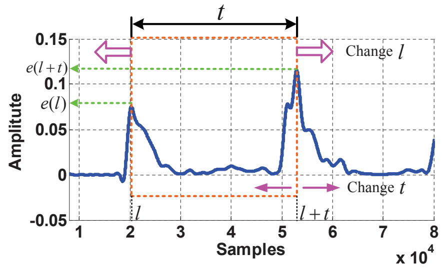


Fig. 4. Illustration of $f(t)$ computation for the envelope $e(l)$ of the breathing sound.

The key difference between the ambient noise and the breathing sound is their stability. This is due to the fact that the amplitude of the ambient noise does not vary significantly within a short period (e.g., a few seconds). We thus can detect frames that only contain ambient noise from a sequence of frames by calculating the variance $v_i = \text{var}(F_i)$ for the i-th frame F_i to capture the stability of the acoustic signal within the frame. However, such variances may vary with different noise profiles. To deal with this, our system further performs the variance normalization to achieve a robust noise detection:

$$\overline{v_i} = \frac{v_i - v_{mean}}{v_{std}} \tag{1}$$

where v_{mean} and v_{std} denote the mean and standard deviation of variance values within an observation window respectively. After normalization, we conduct a statistical study and empirically choose 1 as the threshold to identify the noise frames as follows: if $\overline{v_i} < 1$, frame F_i will be detected as the frame that only contain ambient noise and vice versa.

2) Noise Subtraction: After detecting the noise frames, our system performs noise subtraction by estimating the noise spectral magnitude from the detected frames. Specifically, let $\{\hat{r}(l), 0 \le l \le N-1\}$ be a sequence of acoustic samples after filtering and $\{n(l), 0 \le l \le N-1\}$ be the noise that has been added to the clean acoustic signal $\{r(l), 0 \le l \le N-1\}$. Their relationships thus can be represented as $\{\hat{r}(l) = r(l) + n(l), 0 \le l \le N-1\}$. Taking the Fourier transform, we have:

$$\hat{R}(e^{jw}) = R(e^{jw}) + N(e^{jw}) \tag{2}$$

where $\hat{R}(e^{jw})$, $R(e^{jw})$ and $N(e^{jw})$ denote the Fourier transform of sequence $\hat{r}(l)$, $r(l)$ and $n(l)$, respectively. To estimate the frequency spectrum of the breathing sound $R(e^{jw})$, we can estimate the noise magnitude spectrum $N(e^{jw})$ and then subtract it from the spectrum of the recorded acoustic data $\hat{R}(e^{jw})$:

$$R(e^{jw}) = [\left|\hat{R}(e^{jw})\right| - E(\left|N(e^{jw})\right|)]e^{j\theta_{\hat{R}(e^{jw})}} \tag{3}$$

where the magnitude of the noise is estimated using the average value $E(\left|N(e^{jw})\right|)$ derived from detected noise frames and the phase is estimated by the phase of $\hat{R}(e^{jw})$. Finally, we can obtain the cleaned acoustic signal $\{r(l), 0 \le l \le N-1\}$ after taking the Inverse Fourier transform on $R(e^{jw})$.

B. Breathing Rate Detection

After noise reduction, we first extract the envelope of the acoustic signal. We then utilize the strong correlation relationship between breathing cycles to search for the time length between the breathing cycle pattern to derive the breathing rate.

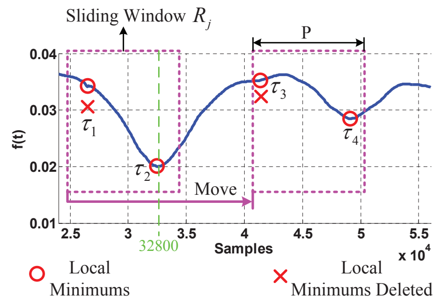


Fig. 5. Illustration of Local Minimum Removal.

1) Envelope Detection: The envelope captures the trend changes of the acoustic signal and thus can be used for the breathing cycle detection. Given N acoustic samples have been equally divided into L frames with $K = 800$ samples each: $F_i = \{r(l), (i-1)K \le l < iK - 1\}, i = 1, ..., L$ with $LK = N$, we first computes the maximum absolute value of the acoustic samples in each frames as $M_i = \max(|F_i|) = \max\{|r(l)|, (i-1)K \le l < iK - 1\}$. We then perform the interpolation to make the length of each sequence consistent. The interpolation allows the sequence of maximum absolute values having the same length as the sequence of the input acoustic samples. In particular, we align the extracted sequence $\{M_i, i = 1, ..., L\}$ to the length N by using the cubic spline interpolation. The sequence after interpolation is considered as the detected envelope and can be represented as:

$$\{e(0), ..., e(N-1)\} = Interpolation\{M_1, ..., M_L\} \tag{4}$$

2) Breathing Rate Identification: The identified envelope $\{e(l), l = 0, ..., N-1\}$ contains the information of the breathing cycles. To identify the breathing rate, we utilize the correlation inherent in the user's breathing cycles. The correlation among breathing cycles allows us to detect a user's breathing rate accurately. This is because the time length between the breathing cycle patterns is resistant to the acoustic sample distortion caused by irregular breathing. To identify this time length, we examine the similarity between acoustic samples as a function of the time lag between them. The basic idea is that the acoustic samples should be highly correlated when the time lag is equal to the breathing period. Such time lag can be then used to identify the breathing rate.

Note that the period of a periodic signal (i.e., the envelope $\{e(l), l = 0, ..., N-1\}$) is defined as the smallest amount of samples it takes to repeat itself. Thus, it always satisfies $e(l) \approx e(l+t)$ if t is equal to the period. We assume the minimum and maximum interval between the breathing cycles under normal conditions have $T_{\min}$ and $T_{\max}$ samples respectively. Given the range of adult's possible breathing rate and the sampling frequency of the microphone, we can determine reasonable values for $T_{\min}$ and $T_{\max}$ as 24000 and 80000 samples, respectively. We then define a function $f(t)$ to measure the similarity between acoustic samples as a function of the time lag t between them:

$$f(t) = \frac{\sum_{l=0}^{N-T_{\max}-1} |e(l) - e(l+t)|}{N - T_{\max}}, T_{\min} \le t \le T_{\max} \tag{5}$$

We illustrate the $f(t)$ computation in Figure 4. When we increase the value of t, the values of $f(t)$ increase and decrease successively, indicating similarity between any pairs of samples in the envelope $\{e(l), l = 0, ..., N-1\}$ with a distance of t: larger values denote the similarity is lower while smaller values denote the similarity is higher. Thus, to identify the breathing rate, we search for a set of W local minimums (i.e., $MinSet = \{\tau_k, 1 \leq k \leq W\}$) from $f(t)$ by varying t from $T_{\min}$ to $T_{\max}$. For each $\tau_k \in MinSet$, it satisfies that $f(\tau_k) < f(t)$ for any $t \in (\tau_k - d, \tau_k + d)$, where d is a pre-defined small distance with $d > 0$. Since the period of a periodic signal is defined as the smallest amount of samples it takes to repeat itself, ideally, the first local minimum τ_1 should therefore correspond to the period of envelope $\{e(l), l = 0, ..., N-1\}$.

However, identifying the τ_1 accurately is challenging because the detected first local minimums could be affected by the noise existed in the acoustic sound. For this reason, we utilize the fact that the consecutive breathing cycles should have a high similarity in the collected acoustic data. Thus, the "true" first local minimum that corresponds to the period of envelope should hold a relatively smaller value than other nearby local minimum points. So it is natural for us to think about using a sliding window to remove the local minimums with relatively larger values. Specifically, given the window length P and consecutive P samples $R_j = \{f(t), j \leq t < j + P\}$ from $f(t)$, for all the local minimums in R_j, we only keep the local minimum which equals to the smallest value of R_j and remove other local minimums from the set $MinSet$. In this work, we empirically set the P as the number of samples a typical breathing cycle has with $P = 40000$ samples and the algorithm of local minimum removal is provided in Algorithm 1. After the removal process, the "true" first local minimum τ_1 which corresponds to the period of envelope $\{e(l), l = 0, ..., N-1\}$ can therefore be detected. And the breathing rate thus can be identified as $1/\tau_1$.

Figure 5 shows an example on how the local minimums are removed in $MinSet$ from a real experiment. In this experiment, the groundtruth of the user's breathing rate is about 14.5 breaths per minute (bpm) (i.e., the breath interval is about 32800 samples). In Figure 5, the detected local minimums in $MinSet$ before local minimum removal are shown as red circles. To conduct the local minimum removal, a window with the length P slides across the sequence of $f(t)$ values: for each window position, only the local minimum point that equals to the smallest value in the window is kept and other local minimums are deleted. The deleted points are shown as red crosses. Before local minimum removal, the detection of first local minimum is not accurate due to τ_1 does not correspond to the true breath interval of the user. After local minimum removal, we find that the previous first local minimum τ_1 has been deleted and τ_2 becomes the new first local minimum, which is detected as the user's breath interval. This value is very close to the user's true breath interval as shown by the green dotted line in Figure 5. This result is encouraging as it indicates that our local minimum removal algorithm can help our system to achieve an accurate and robust breathing rate detection.

Algorithm 1 Local Minimum Removal

INPUT:
$f(t)$; *Pre-defined function*
$MinSet = \{\tau_k, 1 \leq k \leq W\}$; *$W$ local minimums of $f(t)$*
P; *Length of the sliding window*
$T_{\min}$; *The minimum breath interval*
$T_{\max}$; *The maximum breath interval*
PROCEDURES:
for All $j \in [T_{\min}, T_{\max} - P]$ **do**
  $R_j = \{f(t), j \leq t < j + P\}$;
  for All $k \in [1, W]$ **do**
    if $f(\tau_k) \in R_j \& f(\tau_k) > min(R_j)$ **then**
      delete τ_k from $MinSet$
    end if
  end for
end for
Return $MinSet$

C. Event Detection

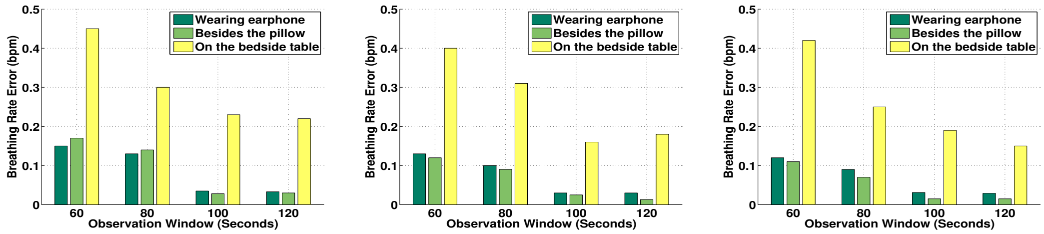
To detect the sleep related events, we first extract the acoustic features (i.e., Mel-frequency cepstral coefficients) from the recorded sound. Based on the extracted features, we utilize Support Vector Machine (SVM) [19] as classifier to detect and identify each sleep event including including snore, cough, turn over and get up.

As shown in [20], the Mel-frequency cepstral coefficients (MFCCs) can characterize the sound's unique characters and are not sensitive to varying acoustic profiles. The MFCCs can be used to distinguish sound frames of four different sleep events (i.e., snore, cough, turn over and get up). Thus, we compute the MFCCs for each frame $F_i = \{r(l), (i-1)K \leq l < iK - 1\}$ that is not detected as noise frame and consider each coefficient within the MFCCs as a feature (i.e., we have 12 features for each F_i). In our classification model, we label certain sleep event's acoustic data as the positive class and all other sleep events' data as the negative class. In particular, to train the SVM classifier for each sleep event, we select U sound frames from this event and labeled as positive instances. We then choose U sound frames from each of the rest events (e.g., the rest is three events), labeled as negative instances. Thus, we have U positive instances and $U \times 3$ negative instances as our training set for each specific sleep event. Training instances (including positive and negative ones) are put together in the training data set to train the SVM classifier. In the sleep event detection phase, the extracted MFCCs obtained from a run-time non-noise frame is input to the event detection model and then SVM classifier outputs a predictive label. If the label is positive, the event detection is a success of this time frame. Otherwise, the label is negative, indicating this time frame does not include the sleep event we are interested.

We find that the detection performance stabilizes when U exceeds 100 instances, which indicates that we need to have at least 10-second sound data from each event for training to create a stable SVM classifier. Thus, unless otherwise specified, we choose $U = 100$ instances in this paper.

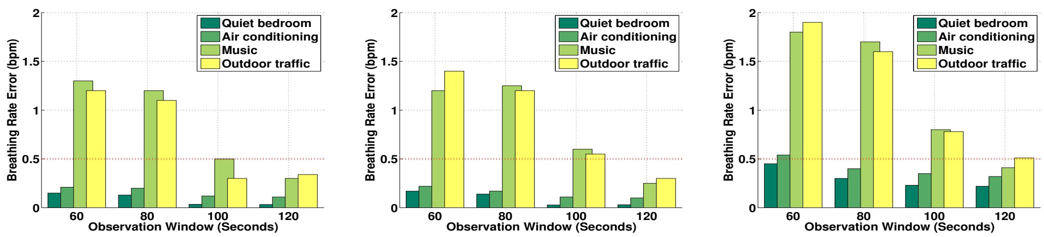
D. Deriving Apnea-hypopnea Index from Breathing Rate

The apnea or hypopnea is a type of sleep disorder which involves episodes of pauses in breathing or the abnormally low respiratory rate [4]. Such sleep disorder may result in a



(a) Microphone only (b) Earbuds + microphone (c) Stereo recording

Fig. 6. Breathing rate detection under different earphone placements and different recording methods.



(a) Wearing earphone (b) Besides the pillow (c) On the bedside table

Fig. 7. Breathing rate detection under different environmental noise.

decreased amount of air movement into the lungs and is also closely linked to some other serious diseases such as obesity, heart disease and diabetes. However, the apnea or hypopnea is hard for an individual to be aware of during sleep [21]. Our fine-grained sleep monitoring system has the capability to monitor the breathing rate, making it possible for the users to detect the apnea or hypopnea.

To quantitative measure the severity of users' apneas and hypopneas, we exploit user's detected to derive the apnea-hypopnea index. This index reports on respiratory events during sleep and it is defined as the average number of apnea or hypopnea events per hour of sleep. In our implementation of index estimation, we adopt the criteria that a time period (i.e, 30 seconds) with apnea or hypopnea is determined by a low detected average breathing rate: if the user's breathing rate is less than the lower bound of adults' normal respiratory rate (i.e., 12 bpm), the label of this time period will be marked as "apnea or hypopnea". Otherwise, the label is marked as "normal", indicating this time period does not include any apnea or hypopnea event. Finally, the apnea-hypopnea index can thus be calculated by dividing the total number of time periods that are marked as "apnea or hypopnea" during the sleep period by the total number of hours of sleep.

V. Performance Evaluation

In this section, we conduct experiments with six subjects over six months time period to evaluate the effectiveness of our fine-grained sleep monitoring system and the sleep apnea monitoring supported by our system. The following subsections detail our experimental methodology and results.

A. Experimental Methodology

We use two iPhone 4 smartphones together with their original earphones that support 8 kHz sampling rate for acoustic data collection. Each iPhone 4 smartphone runs IOS 7 operation system with 512 MB RAM and a 1 GHz Cortex-A8 processor. The acoustic readings are collected when the users are sleeping and then written into a sound file on the smartphone. During the experiments, we let users connect the earphone to smartphones using three implementations of earphone recording presented in Section III-C (i.e., *Microphone only*, *Earbuds + microphone*, and *Stereo recording*) and then place the earphone in three different positions: the participant wears the earphone (i.e., *Wearing earphone*), the participant puts the earphone besides the pillow (i.e., *Besides the pillow*), and on the bedside table (i.e., *On the bedside table*). Such positions are natural choices since many people place earphones together with their smartphones in similar positions during sleep. We conduct our experiments under four representative environments: the quiet bedroom (i.e., *Quiet bedroom*), the relatively noisy bedroom with the air conditioning on (i.e., *Air conditioning*), with the music on (i.e., *Music*), and with the outdoor traffic noise (i.e., *Outdoor traffic*).

We conduct experiments with 6 volunteers (ranging from 23 to 34 years old) over a period of 6 months to evaluate the effectiveness of our system in breathing rate monitoring and sleep events detection. A size of 6 users is also typical for sleep monitoring studies [8], [15]. Unless otherwise stated, the user chooses the microphone on earphone in the recording and places the earphone besides the pillow in a quiet bedroom during sleep.

To obtain the ground truth of the breathing rate, the NEU-LOG Respiration Monitor Logger Sensor [22] is connected to a monitor belt attached to the user's ribcage. The ground truth of the breathing rate is then calculated based on the air pressure changes in the attached monitor belt. Moreover, a lavalier microphone is clipped to the user's collar to collect the high-quality audio clips as the ground truth of sleep events such as snore, cough, get up and turn over. A laptop is connected to the lavalier microphone to store recorded clips. After the data collection, the sleep events are then labeled manually from the audio clips recorded by lavalier microphone.

Several metrics are used to quantify the performance of our

system for breathing rate detection and sleep event detection. For the metrics of sleep event detection, similar in [9], we also consider sleep events with short duration (i.e., snore and cough events) and sleep events with long duration (i.e., turn over and get up event) separately. These metrics are detailed as follows:

- **Breathing rate error:** the difference between the rate detected by our system and the actual rate.
- **True positive rate (for short-duration events):** the ratio of the number of detected positive events to the total number of positive events.
- **False positive rate (for short-duration events):** the ratio of the number of negative events that are mistakenly detected as positive events to the total number of negative events.
- **True positive rate (for long-duration events):** the ratio of the number of the successfully detected time frames associated with positive events to the total number of time frames associated with positive events.
- **False positive rate (for long-duration events):** the ratio of the number of time frames associated with negative events that are mistakenly detected as positive events to the total number of time frames associated with negative events.

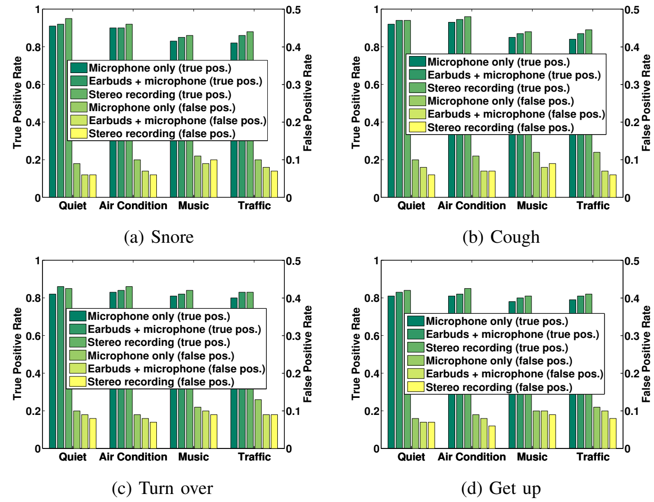
B. Breathing Rate Detection

In the first set of experiment, we evaluate the breathing rate detection under different earphone placements and noisy environments.

Overall Performance. Figure 6 (a) to (c) present the accuracy of breathing rate detection with different lengths of observation window when the earphone is placed in three different positions for three different earphone recording methods. We observe that the breathing rate error is within 0.5 bpm under various position using these three recording methods. As mentioned in [14], the rate errors less than 0.5 bpm is insignificant because only an integer number of breathing rate is reported by current medical devices. This demonstrates that our system is effective in breathing rate detection with different placements of earphone under different recording methods.

Further, we find that the lower breathing rate error is achieved when "Earbuds + microphone" and "Stereo recording" are used. This is because the recording sound can be further enhanced by using earbuds. In addition, we observe that the overall breathing rate error is extremely low (e.g., 0.1) if we place the earphone close to the user, such as wearing earphone and besides the pillow. This is because that shorter distance between the user and the earphone leads to a higher sound intensity at microphone.

Finally, the figure clearly shows that longer observation window results in obvious lower breathing rate error when the window length is shorter than 100 seconds, and this error becomes relatively stable when the window length is longer than 100 seconds. In particular, our system can achieve the breathing rate error of less than 0.05 bpm. While more breathing cycles exist in a longer observation window which captures the breathing rate more accurately, we find that a window length of 100 seconds is enough for our scheme to achieve a high accuracy of breathing rate detection.



(a) Snore

(b) Cough

(c) Turn over

(d) Get up

Fig. 8. Sleep event detection under different environmental noise and different recording methods.

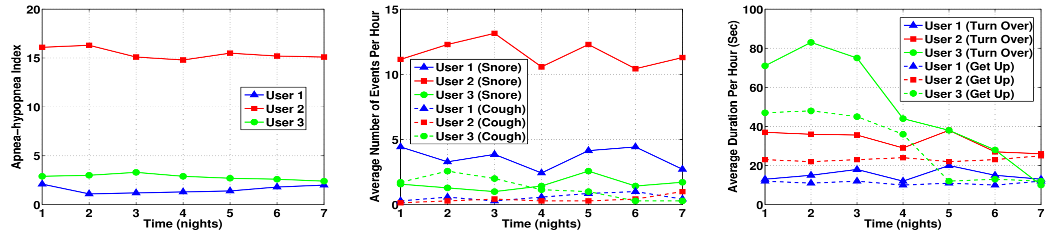
Impact of Environment Noise. We next study the impact of environmental noise on the breathing rate detection. In different environments, the "*Microphone only*" is used for recording and the earphone is placed in three different positions.

Figure 7 (a) to (c) present the accuracy of breathing rate detection under different environments. We observe that the overall breathing rate error remains less than 0.5 bpm (as the red dotted line shown in the figure) across all environments and earphone placements when the window length is longer than 100 seconds. This demonstrates that our sleep monitoring system can achieve a satisfactory breathing rate estimation with 100 seconds observation window. Further, this figure clearly shows that similar detection accuracy is achieved in both the quiet bedroom and the room with the air conditioning on, indicating that our system is robust to the noise.

C. Sleep Event Detection

Next, we evaluate our system by investigating the accuracy of detecting sleep related events under various environments with three earphone recording methods. The legends "true pos." and "false pos." in Figure 8 denote the detection rate and false positive rate, respectively. For each experiment, the earphone is placed besides the pillow.

Figure 8 (a) to (d) depict the true positive and false positive rate under various noisy environment. We observe that better performance can be achieved when "Earbuds + microphone" and "Stereo recording" are used. This is because the earbuds can improve the recorded sound quality. Further, we find that the true positive rate increases and false positive rate decreases when we change from noisy environment to quiet bedroom. This is because in a relatively quiet environment, it is easier to capture and identify the sound of different sleep events from the noise. Overall, we observe that that our system can achieve over 80 % true positive rate with less than 10 % false positive rate in all scenarios. This shows that our system is effective in detecting sleep related evens and robust across different environments.



(a) Apnea-hypopnea index (b) Snore and cough events (c) Turn over and get up events

Fig. 9. Case study: the apnea-hypopnea index and sleep events of different users during the 7-night experiment.

D. Case Study: Sleep Apnea Monitoring

Finally, we present the results of a one-week case study by involving three users to evaluate the performance of the fine-grained sleep quality monitoring. Specifically, we calculate the Apnea-hypopnea Index based on the detected breathing rate for sleep apnea monitoring. One user (i.e., user 2) has sleep hypopnea and another user (i.e., user 3) catches a cold during the period of study. User 2 who has sleep hyopnea occasionally has an abnormal low breathing rate for a duration of about one minute during sleep. User 3 who catches a cold coughs more frequently than others. In this study, the earphone is placed besides the pillow and the earphone is used for recording the sound.

Figure 9 presents the apnea-hypopnea index and events detected by our system in each night in one week. We observe that the apnea-hypopnea index of user 2 is higher than other users. This is due to that user 2 has periodical low breathing rate during sleep. It illustrates that our system can capture the user's breathing disorders accurately during sleep. Another observation is that more cough events are detected for user 3 during the first 5 days of the study. This is because that user 3 catches a cold at the beginning of the study and recover gradually. The captured number of cough events thus decrease gradually as user 3 recovers. In addition, the figure also demonstrates that users (i.e., user 2 and 3) with discomfort (i.e., cough or hypopnea) are likely to turn over or get up during sleep. This is because: 1) the cough or hypopnea events are usually followed by the user' body movements; and 2) frequent cough makes the user hard to fall asleep. These observations strongly confirm the effectiveness of using our system for supporting healthcare related application, such as sleep apnea monitoring.

VI. Conclusion

In this work, we focus on achieving fine-grained sleep monitoring by leveraging smartphones. We propose a practical system that has the capability to monitor an individual's breathing rate as well as sleep events using off-the-shelf smartphones. In particular, our system employs the readily-available earphone for smartphones to capture the breathing sound and measure the breathing rate. Our noise detection and subtraction scheme can reduce the impact of background noise while preserving the features present in the breathing sound for breathing rate detection. Furthermore, our system exploits the correlation relationship inherent in a user's breathing cycles to identify breathing rate accurately based on the signal envelope detection. Through extensive experiments over six months time period, we show that the breathing rate monitoring is highly accurate and robust under various environments. This strongly indicates the feasibility of using the smartphone and its earphone to perform continuous and noninvasive fine-grained sleep monitoring. We further demonstrate that our system can be used to support healthcare related applications through a case study of sleep apnea monitoring.

References

[1] J. M. Parish, "Sleep-related problems in common medical conditions," *Chest Journal*, 2009.

[2] J. E. Shaw, N. M. Punjabib, J. P. Wildingc, K. G. M. Albertid, and P. Z. Zimmeta, "Sleep-disordered breathing and type 2 diabetes: A report from the international diabetes federation taskforce on epidemiology and prevention," *Diabetes research and clinical practice*, 2008.

[3] H. Dincer and W. O'Neill, "Deleterious effects of sleepdisordered breathing on the heart and vascular system," *Respiration*, 2006.

[4] "Sleep apnea: What is sleep apnea?" *NHLBI: Health Information for the Public. U.S. Department of Health and Human Services.*, 2010.

[5] P. X. Braun, C. F. Gmachl, and R. A. Dweik, "Bridging the collaborative gap: Realizing the clinical potential of breath analysis for disease diagnosis and monitoring tutorial," *IEEE Sensors Journal*, 2012.

[6] G. S. Chung and et al, "Rem sleep classification with respiration rates," in *Proceedings of IEEE ITAB*, 2007.

[7] C. A. Kushida and et al, "Practice parameters for the indications for polysomnography and related procedures: an update for 2005," *Sleep*, 2005.

[8] "Sleep tracker," http://www.sleeptracker.com/, 2014.

[9] T. Hao, G. Xing, and G. Zhou, "isleep: Unobtrusive sleep quality monitoring using smartphones," in *Proceedings of ACM Sensys*, 2013.

[10] A. Yadollahi and Z. Moussavi, "Comparison of flow-sound relationship for different features of tracheal sound," in *Proceedings of IEEE EMBC*, 2008.

[11] "PRWeb News," http://www.prweb.com/releases/2013/1/, 2013.

[12] E. van Dijk, T. I. M. Hilgenkamp, H. M. Evenhuis, and M. A. Echteld, "Exploring the use of actigraphy to investigate sleep problems in older people with intellectual disability," *Journal of Intellectual Disability Research*, 2012.

[13] F. Takatori, S. Yamamori, M. Inoue, S. Abe, and K. Miyasaka, "A novel mainstream capnometer system for non-intubated pediatric patients requiring oxygen administration," in *IEEE Eng Med Biol Soc*, 2011.

[14] N. Patwari, J. Wilson, S. Ananthanarayanan, S. K. Kasera, and D. R. Westenskow, "Monitoring breathing via signal strength in wireless networks," *IEEE Transactions on Mobile Computing*, 2013.

[15] Z. Chen and et al, "Unobtrusive sleep monitoring using smartphones," in *Proceedings of PervasiveHealth*, 2013.

[16] "Respiratory Rate," http://appcrawlr.com/ios/respiratory-rate, 2014.

[17] "AR-4i," http://www.fostexinternational.com/docs/products/AR-4i.shtml, 2014.

[18] N. Gavriely, M. Nissan, A. H. Rubin, and D. W. Cugell, "Spectral characteristics of normal breath sounds," *J Appl Physiol*, 1981.

[19] T. Plötz, N. Y. Hammerla, A. Rozga, A. Reavis, N. Call, and G. D. Abowd, "Automatic assessment of problem behavior in individuals with developmental disabilities," in *Proceedings of ACM UbiComp*, 2012.

[20] A. Zolnay, R. Schluter, and H. Ney, "Acoustic feature combination for robust speech recognition," in *Proceedings of IEEE ICASSP*, 2005.

[21] "Sleep apnea: Key points," *NHLBI: Health Information for the Public. U.S. Department of Health and Human Services.*, 2014.

[22] "NEULOG Respiration Monitor Logger Sensor," http://www.neulog.com/, 2014.